

Formát dat přijímaných digitální knihovnou

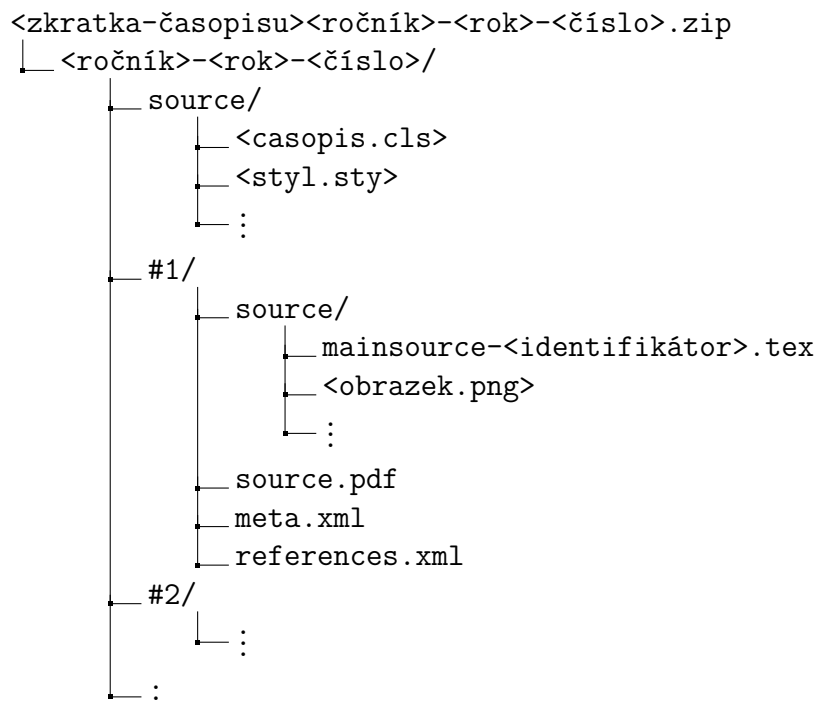
\$Revision: 2626\$

Obsah

Obsah	ii
1 Adresářová struktura ZIP archivu s předávanými daty	1
1.1 Soubor <code><zkratka-časopisu><ročník>-<rok>-<číslo>.zip</code>	1
1.2 Adresář <code><ročník>-<rok>-<číslo></code>	1
1.3 Adresář <code>source</code> v kořenovém adresáři archivu	2
1.4 Adresáře <code>#<pořadové-číslo></code> v kořenovém adresáři archivu	2
1.5 Adresář <code>source</code> v podadresáři článku	2
1.6 Soubor <code>mainsource-<identifikátor>.tex</code>	3
1.7 Soubor <code>source.pdf</code>	3
1.8 Soubor <code>meta.xml</code>	3
1.9 Soubor <code>references.xml</code>	3
2 Formát metadat	4
2.1 Článeková metadata – soubor <code>meta.xml</code>	5
2.2 Metadata referencí – soubor <code>references.xml</code>	7
2.2.1 Základní značkování	7
2.2.2 Podrobné značkování	9

1 Adresářová struktura ZIP archivu s předávanými daty

Data každého nového čísla časopisu pro digitální knihovnu jsou předávána ve formě ZIP archivu. Adresářová struktura tohoto archivu je následující:



1.1 Soubor <zkratka-časopisu><ročník>-<rok>-<číslo>.zip

Vhodně pojmenovaný archiv s metadaty pro digitální knihovnu.

Příklad: ao049-2010-3.zip

1.2 Adresář <ročník>-<rok>-<číslo>

Vhodně pojmenovaný kořenový adresář archivu s metadaty pro digitální knihovnu. Ročník je zadán přesně třemi číslicemi (doplňuje se nulami zleva). Adresář obsahuje právě jedno číslo časopisu.

Příklad: 049-2010-3

1.3 Adresář source v kořenovém adresáři archivu

Adresář obsahuje *všechny* zdrojové soubory nutné pro překlad daného čísla časopisu, které přímo nenáleží k některému konkrétnímu článku čísla. Vkládají se pouze soubory specifické a upravené pro konkrétní časopis, není vhodné vkládat soubory obvykle obsažené v distribuci T_EXu (např. `article.cls`, `hyperref.sty` apod.), pokud nebyly oproti distribuční verzi modifikovány.

Příklad: Adresář obsahuje redakční soubory `casopis.cls` a `styl.sty` s definicemi layoutu časopisu a soubor `logo.png` s logem tištěným na obálku. Dále pak soubory `obsah.tex` s ručně vytvořeným obsahem a `obalka.tex` s obálkou daného čísla.

1.4 Adresáře #<pořadové-číslo> v kořenovém adresáři archivu

Sada adresářů s jednotlivými články čísla časopisu. Každý adresář má pořadové číslo (číslováno od 1), které *odpovídá fyzickému pořadí článků* v daném čísle časopisu. Mezi články se v tom případě počítají (a tudíž dostanou přiděleny své adresáře #<pořadové-číslo>) také předmluvy redakce, redakční oznámení apod.

Příklad: #5

1.5 Adresář source v podadresáři článku

Adresář obsahuje *všechny* zdrojové soubory, které jsou (ve spojení se soubory z adresáře `source` v kořenovém adresáři archivu, vizte oddíl 1.3) nutné pro překlad daného čísla časopisu a přímo náležejí ke článku umístěném v tomto podadresáři. Povinně je přítomen zejména hlavní zdrojový soubor článku `mainsource-<identifikátor>.tex` (vizte oddíl 1.6 na následující straně), dále pak případné další soubory ke článku náležející. Vhodné je také dokumenty překládat s nastavením `\listfiles` a následně přiložit log soubor, který vznikl při překladu článku.

Příklad: Adresář obsahuje hlavní zdrojový soubor `mainsource-novak-praha.tex`, T_EXové soubory `novak-praha-cast1.tex` až `novak-praha-cast4.tex` s kapitoly článku načítanými z hlavního zdrojového souboru, sadu obrázků `obr1.png` až `obr14.png`, které jsou v článku používány, a log překladu článku `mainsource-novak-praha.log`.

1.6 Soubor `mainsource-<identifikátor>.tex`

Hlavní zdrojový soubor článku. K zajištění snadné orientace při případném výskytu více \TeX ových zdrojových textů (např. jednotlivé kapitoly článku v samostatných \TeX ových souborech, které jsou následně načítány v hlavním zdrojovém souboru) obsahuje jeho jméno předponu „mainsource“ následovanou spojovníkem („-“) a libovolným identifikátorem redakce (např. původním názvem zdrojového textu dodaného redakci autorem). Předponu je nutno uvádět vždy, i když je zdrojový text jediným souborem v adresáři `source` v podadresáři článku (vizte oddíl 1.5 na předchozí straně).

Příklad: `mainsource-novak-praha.tex`

1.7 Soubor `source.pdf`

Finální přeložená podoba daného článku ve formátu PDF přesně tak, jak byla vytištěna ve fyzické podobě daného čísla časopisu. Jelikož je tento dokument určen pro dlouhodobou archivaci, je vhodné, aby odpovídal normě PDF/A (ISO 19005-1:2005), anebo se jí alespoň maximálně přibližoval (zejména jsou všechny použité fonty, obrázky a jiné komponenty fyzicky vloženy v dokumentu, soubor není šifrovaný, nevyužívá JavaScript atd.).

Příklad: `source.pdf`

1.8 Soubor `meta.xml`

Metadata ve formátu XML popisující daný článek. Pro popis formátu článkových metadat vizte sekci 2.1 na straně 5.

Příklad: `meta.xml`

1.9 Soubor `references.xml`

Metadata ve formátu XML popisující seznam literatury daného článku. Pro popis formátu metadat referencí vizte sekci 2.2 na straně 7. Pokud článek neobsahuje seznam literatury (jedná se např. o redakční oznámení apod.), není tento soubor v daném podadresáři přítomen.

Příklad: `references.xml`

2 Formát metadat

Metadata se předávají ve formátu XML v kódování UTF-8. Řídící sekvence \TeX u se plně expandují a výsledek expanze se na výstup předává jako UTF-8 text, případně se \TeX ové sekvence odstraňují bez jakéhokoliv viditelného efektu na výstup tam, kde nemají v metadatech smysl (explicitní zlom stránky nebo řádku, penalty apod.).

Výjimku tvoří zápis matematiky, který se uvádí v původní neexpandované podobě, neboť tento zápis vzorců je pro matematiky nejpřirozenější.¹ Každý matematický zápis v metadatech se ohraničuje jednou z těchto dvojic oddělovačů:

`$...$` Vyznačuje vzorce uvnitř textu, které byly v originálním zdrojovém textu označeny dvojicí `$` a `$`, `\(` a `\)`, či `\begin{math}` a `\end{math}`.

`\[...\]` Vyznačuje vysazené vzorce v originálním zdrojovém textu označené dvojicí `$$` a `$$`, `\[` a `\]`, či `\begin{displaymath}` a `\end{displaymath}`.

`\begin{equation}...\end{equation}` Vyznačuje vysazené vzorce v originálním zdrojovém textu označené dvojicí `\begin{equation}` a `\end{equation}`.

Obdobným způsobem vyznačujeme také prostředí `align`, `align*`, `alignat`, `alignat*`, `equation*`, `eqnarray`, `eqnarray*`, `flalign`, `flalign*`, `gather`, `gather*`, `multline`, či `multline*`.

Upozornění: Všechny příklady jsou určeny pouze k předvedení technického řešení. Po obsahové stránce nedávají smysl.

Příklad:

Zdrojový text

```
%%%%%%%%%
\def\ALX{Alexanderov}
\def\adtv#1{$\csname#1\endcsname$-additive}
%%%%%%%%%
Ond\v rej Nov\'{a}k\newline
In case  $E_{n}^{*}$  separates the points of  $E$ , an  $\text{\ALX}$ 's
```

¹Cílem je, aby zůstala zachována matematikovi srozumitelná podoba vzorců. V případě použití specifických maker (u nichž obecně nelze předpokládat, že čtenář zná jejich definici) je proto vhodnější tato makra expandovat do takové úrovně, která bude srozumitelná běžnému uživateli \LaTeX u/ $\mathcal{A}\mathcal{M}\mathcal{S}$ - \LaTeX u.

```

type theorem is
proved for
a sequence of
\adtv{\sigma} measures.% TODO
\vspace{5em}
\[ e =
    mc^2 \]

```

se převede na jediný řádek v kódování UTF-8:

```

Ondřej Novák In case  $E_n^*$  separates the points of  $E$ ,
an Alexanderov's type theorem is proved for a sequence of
 $\sigma$ -additive measures.  $[e = mc^2]$ 

```

Aby mohla metadata sloužit jako samostatný celek, je nutné potlačit v nich křížové odkazy. Pakliže některá položka metadat (typicky abstrakt článku) obsahuje citační odkaz, je nutné tento odkaz nahradit přímo citačním záznamem.

Příklad:

Část zdrojového textu `in~\cite{1}.`, která je ve vysázeném dokumentu vytištěna jako `in [1].`, se v metadatech objeví jako UTF-8 kódovaný řetězec:

```

in [Agarwal, R. P.; O'Regan, D.: Existence of three solutions to
integral and discrete equations via the Leggett-Williams fixed
point theorem. Rocky Mountain J. Math. 31 (2001), 23-35.].

```

Oddíly 2.1 a 2.2 obsahují neformální popis XML jazyka metadat. Formální specifikace v podobě XML schématu je součástí on-line validační služby metadat.

2.1 Článeková metadata – soubor `meta.xml`

Soubor `meta.xml` obsahuje metadata popisující článek jako celek. Kořenový element XML dokumentu má jméno `article` a obsahuje následující potomky:

title Název článku. Obsahuje atribut `lang` specifikující zkratkou jazyk, ve kterém je název uveden.

author Jméno autora článku ve formátu „Příjmení, Jméno“. Obsahuje atribut `order` s pořadovým číslem (číslováno od 1) jména autora tak, jak bylo umístěno v článku.

language Zkratkou uvedený jazyk, ve kterém je článek napsán.

keyword Právě jedno klíčové slovo. Obsahuje atribut `lang` specifikující zkratkou jazyk, ve kterém je klíčové slovo zapsáno.

summary Abstrakt článku. Obsahuje atribut `lang` specifikující zkratkou jazyk, ve kterém je abstrakt napsán.

lang_summary Zkratkou uvedený jazyk, ve kterém je v článku zapsán (hlavní) abstrakt.

msc Právě jeden Mathematical Subject Classification kód.

category Typ článku. Nabývá následujících hodnot: „math“ – odborný matematický článek, „review“ – odborné recenze knih a článků, „history“ – historie matematiky a matematiků, bibliografie, „editorial“ – redakční článek, „news“ – novinky a oznámení.

range_pages Rozsah stránek ve formátu „<číslo-první-strany>-<číslo-poslední-strany>“ tak, jak je vytištěn v článku.

U výše uvedených elementů uvádějících jazyk zkratkou se užívá kódů dle ISO 639-2:1998 B: „cze“ – čeština, „eng“ – angličtina, „fre“ – francouzština, „ger“ – němčina, „rus“ – ruština, „slo“ – slovenština atd.

Příklad:

```
<?xml version="1.0" encoding="UTF-8"?>
<article>
  <title lang="eng">Stochastic approximation methods</title>
  <author order="1">Novák, Ondřej</author>
  <author order="2">Khurana, Surjit Singh</author>
  <language>eng</language>
  <keyword lang="eng">order convergence</keyword>
  <keyword lang="eng">$p$-Laplacian</keyword>
  <keyword lang="eng">Alexandrov's theorem</keyword>
  <summary lang="eng">Let  $X$  be a completely regular  $T_1$  space,  $E$ 
a boundedly complete vector lattice,  $C(X)$   $C_b(X)$  the space of all
(all, bounded), real-valued continuous functions on  $X$ . In order convergence,
we consider  $E$ -valued, order-bounded,  $\sigma$ -additive,  $\tau$ -additive,
and tight measures on  $X$  and prove some order-theoretic and topological
properties of these measures. Also for an order-bounded,  $E$ -valued (for some
special  $E$ ) linear map on  $C(X)$ , a measure representation result is proved.
In case  $E_n^*$  separates the points of  $E$ , an Alexanderov's type theorem
```



```

is proved for a sequence of  $\sigma$ -additive measures.</summary>
<lang_summary>eng</lang_summary>
<msc>28A33</msc>
<msc>28B15</msc>
<msc>46G10</msc>
<msc>46B42</msc>
<category>math</category>
<range_pages>307-316</range_pages>
</article>

```

Pro každý dokument se vyplňují všechna pole, která jsou pro něj k dispozici. U odborného matematického článku by to zpravidla měla být všechna z uvedeného seznamu, u „ne-článků“ (redakčních oznámení, recenzí apod.) některá pole existovat nemusí (např. chybí explicitně uvedení autoři, abstrakt apod.), pak se v `meta.xml` odpovídající element neuvádí.

Z hlediska mezinárodního využití metadat je vhodné, aby elementy `title`, `keyword` a `summary` byly uvedeny navíc také v anglickém jazyce, pokud je jejich původní jazyk odlišný.

2.2 Metadata referencí – soubor `references.xml`

K zajištění dobrých možností prohledávání, třídění a kvalitního provázání článků nejen v rámci digitální knihovny, ale také do světových citačních databází, je velmi důležité získat kvalitní metadata popisující seznamy referencí uvedených v jednotlivých článcích. Ke každému článku obsahujícímu seznam použité literatury musí být proto vytvořen také soubor `references.xml` s XML metadaty seznam citací popisujících.

Naprosto precizní označkování všech citačních odkazů nemusí být vždy snadné, je proto stanoveno základní jednoduché značkování, které však *musí* být u všech záznamů, kde je to možné, doplněno o další strukturní informace.

2.2.1 Základní značkování

Kořenový element XML dokumentu má jméno `references` a obsahuje následující potomky:

reference Element uzavírající v sobě informace o jedné konkrétní položce ze seznamu literatury. Obsahuje atribut `id` s pořadovým číslem (číslováno od 1) fyzického umístění reference v seznamu citované literatury. Dále uvedené elementy jsou potomky každého výskytu tohoto elementu.

prefix Prefix položky seznamu literatury tak, jak je vytištěn ve fyzické podobě dokumentu.

title Název položky seznamu literatury.

authors Element uzavírající seznam všech autorů dané položky seznamu literatury. Jeho potomky jsou elementy **author** (jeden pro každého autora dané položky) obsahující jména ve formátu „Příjmení, Jméno“.

suffix Element obsahující nijak nestrukturovaný celý zbytek citačního záznamu tak, jak je vytištěn ve fyzické podobě dokumentu.

Příklad:

Dvě položky literatury

- [1] R. P. Agarwal; D. O'Regan: Existence of three solutions to integral and discrete equations via the Leggett-Williams fixed point theorem. Rocky Mountain J. Math. 31 (2001), 23-35.
- [2] R. P. Agarwal; D. O'Regan; P. J. Y. Wong: Positive Solutions of Differential, Difference and Integral Equations. Kluwer Academic Publishers, Dordrecht, 1999.

se převedou na:

```
<?xml version="1.0" encoding="UTF-8"?>
<references>
  <reference id="1">
    <prefix>[1]</prefix>
    <title>Existence of three solutions to integral and discrete equations via
the Leggett-Williams fixed point theorem</title>
    <authors>
      <author>Agarwal, R. P.</author>
      <author>O'Regan, D.</author>
    </authors>
    <suffix>. Rocky Mountain J. Math. 31 (2001), 23-35.</suffix>
  </reference>
  <reference id="2">
    <prefix>[2]</prefix>
    <title>Positive Solutions of Differential, Difference and Integral
Equations</title>
    <authors>
```

```

    <author>Agarwal, R. P.</author>
    <author>O'Regan, D.</author>
    <author>Wong, P. J. Y.</author>
  </authors>
  <suffix>. Kluwer Academic Publishers, Dordrecht, 1999.</suffix>
</reference>
</references>

```

Výše uvedený příklad minimálního značkování není dostatečný, neboť pole **suffix** obsahuje jasné odlišitelné další prvky. Je proto nutné použít rozšířeného značkování (vizte oddíl 2.2.2).

2.2.2 Podrobné značkování

Podrobné značkování je *rozšířením* základního značkování položek seznamů literatury článku popsaného v oddíle 2.2.1 na straně 7. Vychází z položek bibliografické databáze BibTeXu a *přidává* další elementy popisující strukturu elementu **suffix** ze základního značkování. Všechny elementy popsané v oddíle 2.2.1 *zůstávají zachovány v uvedené podobě*, a to včetně elementu **suffix**. Položky podrobného značkování jsou:

address Adresa (např. vydavatelství).

booktitle Název knihy, jejíž část je citována.

chapter Číslo kapitoly (oddílu . . .), která je citována.

edition Vydání citovaného dokumentu.

editors Element uzavírající seznam všech editorů dané položky v seznamu literatury. Jeho potomky jsou elementy **editor** (jeden pro každého autora dané položky) obsahující jména ve formátu „Příjmení, Jméno“.

howpublished Způsob vydání dokumentu.

institution Instituce, která podpořila vydání dokumentu.

ISBN ISBN knihy.

ISSN ISSN periodika.

journal Název / zkratka názvu časopisu.

month Měsíc, ve kterém byla/bude práce vydána.

note Poznámka.

number Číslo vydání časopisu.

organization Organizace sponzorující konferenci.

pages Rozsah citovaných stránek ve formátu „<číslo-první-strany>-<číslo-poslední-strany>“.

publisher Jméno vydavatele.

school Jméno školy, kde byla práce napsána.

series Název knižní řady nebo kolekce.

type Typ technické zprávy, např. „Research Note“.

URL URL citovaného dokumentu.

volume Ročník časopisu.

year Rok vydání/napsání díla.

Příklad:

Dvě položky literatury

- [1] R. P. Agarwal; D. O'Regan: Existence of three solutions to integral and discrete equations via the Leggett-Williams fixed point theorem. Rocky Mountain J. Math. 31 (2001), 23-35.
- [2] R. P. Agarwal; D. O'Regan; P. J. Y. Wong: Positive Solutions of Differential, Difference and Integral Equations. Kluwer Academic Publishers, Dordrecht, 1999.

se převedou na:

```
<?xml version="1.0" encoding="UTF-8"?>
<references>
  <reference id="1">
    <prefix>[1]</prefix>
    <title>Existence of three solutions to integral and discrete equations via
the Leggett-Williams fixed point theorem</title>
```

```

<authors>
  <author>Agarwal, R. P.</author>
  <author>O'Regan, D.</author>
</authors>
<journal>Rocky Mountain J. Math.</journal>
<volume>31</volume>
<year>2001</year>
<pages>23-35</pages>
<suffix>. Rocky Mountain J. Math. 31 (2001), 23-35.</suffix>
</reference>
<reference id="2">
  <prefix>[2]</prefix>
  <title>Positive Solutions of Differential, Difference and Integral
  Equations</title>
  <authors>
    <author>Agarwal, R. P.</author>
    <author>O'Regan, D.</author>
    <author>Wong, P. J. Y.</author>
  </authors>
  <publisher>Kluwer Academic Publishers</publisher>
  <address>Dordrecht</address>
  <year>1999</year>
  <suffix>. Kluwer Academic Publishers, Dordrecht, 1999.</suffix>
</reference>
</references>

```